

DINAMelt web server for nucleic acid melting prediction

Nicholas R. Markham¹ and Michael Zuker^{2,*}

¹Department of Computer Science and ²Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180-3590, USA

Received February 14, 2005; Revised April 18, 2005; Accepted May 5, 2005

ABSTRACT

The DINAMelt web server simulates the melting of one or two single-stranded nucleic acids in solution. The goal is to predict not just a melting temperature for a hybridized pair of nucleic acids, but entire equilibrium melting profiles as a function of temperature. The two molecules are not required to be complementary, nor must the two strand concentrations be equal. Competition among different molecular species is automatically taken into account. Calculations consider not only the heterodimer, but also the two possible homodimers, as well as the folding of each single-stranded molecule. For each of these five molecular species, free energies are computed by summing Boltzmann factors over every possible hybridized or folded state. For temperatures within a user-specified range, calculations predict species mole fractions together with the free energy, enthalpy, entropy and heat capacity of the ensemble. Ultraviolet (UV) absorbance at 260 nm is simulated using published extinction coefficients and computed base pair probabilities. All results are available as text files and plots are provided for species concentrations, heat capacity and UV absorbance versus temperature. This server is connected to an active research program and should evolve as new theory and software are developed. The server URL is <http://www.bioinfo.rpi.edu/applications/hybrid/>.

INTRODUCTION

The accurate prediction of melting temperatures for DNA or RNA molecules is important in many biotechnology applications. These include the design of gene probes (1) or other oligonucleotides for use on microarrays, where one of each hybridized pair is immobilized, as well as molecular beacons

(2) or PCR primer design, where both molecules are in solution. The number of applications is very large, and there is a great demand for such calculations.

The most common method in use today for predicting melting temperatures for dimers or for single-stranded, folded monomers is based on a two-state model. Two molecules, A and B, are either hybridized or they are not. The non-hybridized ‘random coil state’ for each molecule is treated as a single reference state. It is usually assumed that A and B are perfectly complementary, so that the hybridized state is obvious. Sometimes, one or more mismatches are permitted in the duplex, including G·T or G·U wobble pairs. In the case of a single, folded molecule, a simple stem–loop structure is assumed. The free energy, enthalpy and entropy changes associated with the transition from ‘hybridized at temperature T ’ to random coil are denoted by ΔG , ΔH and ΔS , respectively. They are related by the equation $\Delta G = \Delta H - T\Delta S$. Both ΔG and ΔH are computed using published nearest neighbor coefficients. We use the ‘unified’ parameters of SantaLucia (3) for DNA and the ‘Turner lab’ parameters for RNA (4).

The melting temperature, T_m °K, for a simple stem–loop structure is computed as $T_m = 1000 \times \Delta H / \Delta S$. The factor of 1000 converts from e.u. (entropy units) to kcal/mol/K. For a dimer, the strand concentrations (mol/l, M) must be considered. If $[A_0]$ and $[B_0]$ are the strand concentrations of A and B, respectively, then the total strand concentration, C_t , is $[A_0] + [B_0]$. The usual assumption is that $[A_0] = [B_0] = C_t/2$. In this case

$$T_m = \frac{1000 \times \Delta H}{\Delta S + R \ln (C_t/f)},$$

where R is the universal gas constant and $f = 4$. For homodimer melting, $[A_0] = C_t$ and the same equation holds with $f = 1$. These computations derive T_m as the temperature at which half of the molecules are folded (stem–loop melting) or dimerized (dimer melting).

The DINAMelt web server addresses the broader challenge of combining up-to-date thermodynamic parameters with appropriate algorithms that compute more than just melting

*To whom correspondence should be addressed. Tel: +1 518 276 6902; Fax: +1 518 276 4824; Email: zukerm@rpi.edu

temperatures. It computes ultraviolet (UV) absorbance, heat capacity (C_p) and concentrations of various dimer and monomer species as a function of temperature. The computed melting profiles can be compared directly with measured data. Heat capacity can be measured using differential scanning calorimetry (DSC) and species mole fractions can be obtained from isothermal titration calorimetry.

Our methods are more general than existing ones in a number of ways.

- (i) The two strands, A and B, are not required to be complementary. A partition function is computed that considers all possible hybridizations or foldings and weights them by their Boltzmann factors.
- (ii) The strand concentrations, $[A_0]$ and $[B_0]$, need not be equal. They can differ by many orders of magnitude.
- (iii) Competition between folding and dimerization is automatically considered for both A and B. Similarly, competition among three dimerized states is taken into account by default. These dimerized states are the usual homodimer, AB, together with the two homodimers, AA and BB.
- (iv) An internal energy term is added to account for the base stacking that is present in single-stranded, unfolded molecules.

It is important to recognize the underlying assumptions made by the DINAMelt server. The simulations are for molecules in solution and the system is assumed to be at thermodynamic equilibrium at each temperature. Predictions made for PCR primers, for example, can be misleading if kinetics are dominant. Hybridization on microarrays is complicated by the fact that one of each hybridizing pair is immobilized. It is difficult to compute the effective 'solution concentration' for such molecules, and diffusion may be an important factor in slowing the equilibration time.

METHODS

The DINAMelt web server uses the methodology described by Dimitrov and Zuker (5). The original software has been completely replaced by a new, integrated collection of programs. The current name for this package is *hybrid* and it is available for download from the DINAMelt website.

Partition functions, Z_x , are computed for each of the five molecular species ($X = A, \dots, AB$) over a range of temperatures, yielding Gibbs free energies of the form $-RT \ln Z_x$. The resulting equilibrium constants are used to derive the concentrations of each species. The species free energies and concentrations are then combined to compute the ensemble free energy. Heat capacities are derived by numerical differentiation of the free energy profiles with respect to temperature.

The partition function calculations also produce base pair probabilities for each species, from which the probabilities that individual bases or dimers are single-stranded can be derived. Finally, computed probabilities and species mole fractions are combined with published extinction coefficients (6) to yield UV absorbance predictions.

A number of different melting temperatures are computed. $T_m(c)$ is defined as the temperature at which the total concentration of all dimers is half of its maximum value at low

temperature. This temperature cannot, in general, be defined as the temperature at which $[AB] = C_t/4$. If $[A_0] < [B_0]$, the excess B will be single-stranded at low temperature unless B hybridizes very well with itself. Even when $[A_0] = [B_0]$, hybridization may be incomplete at low temperature if A and B are poor complements, especially if A or B folds into stable stem-loop structures.

For heat capacity, $T_m(C_p)$ is a temperature at which $\partial C_p / \partial T = 0$. For non-cooperative melting, there may be two or more distinct peaks, leading to two or more values for $T_m(C_p)$. In such cases, the largest computed $T_m(C_p)$ is considered to be the 'melting temperature'.

The server computes two different melting temperatures based on UV absorbance. Inflection points on the profile define $T_m(\text{Ext1})$. As with $T_m(C_p)$, multiple values may be computed. The second computation defines $T_m(\text{Ext2})$ as the midpoint between the minimum computed absorbance and the maximum possible absorbance. Absorbance at low temperatures may be above the zero baseline, even if dimerization is 100% and both strands have the same length. This can happen if A and B are not perfectly complementary so that hybridized states include single-stranded bulges, interior loops and dangling ends that all absorb radiation. It is usual to observe $T_m(\text{Ext1}) - T_m(\text{Ext2}) \approx 1^\circ\text{K}$.

The output section below contains further details on the current presentation of computed melting temperatures. The user should note that text files containing all the predicted T_m s can be downloaded from the server.

A collaboration with IDT (Integrated DNA Technologies, Inc., Coralville, IA) has given us access to DSC melting profiles for several hundred complementary deoxyoligonucleotide pairs. Although some of the melting temperatures have been published (7), the profiles themselves have not. We observed that computed enthalpies are $\sim 10\%$ too small in magnitude compared with measured ones. This systematic error lead us to conclude that the SantaLucia base pair stacking enthalpies did not account for the total enthalpy change. As suggested by Dimitrov and Zuker (5), we attributed the 'missing' enthalpy to an internal energy derived from base stacking in unfolded, single-stranded species. A simple extension to the model was implemented. Single-stranded unfolded molecules are either in the usual random coil reference state or else in a new 'structured state'. The enthalpy and entropy changes between these two states are ΔH_{ss} and ΔS_{ss} , respectively. The 'Advanced Form' subsection of this article describes how these parameters are chosen. A complete description of this correction, together with supporting data, will be published elsewhere.

SERVER CONTENT AND ORGANIZATION

Input

The default form allows the user to submit a simple job with two sequences. (There are also additional forms for jobs with more complicated options or with only one sequence, which will be discussed below.) The user fills in several fields, most of which have certain constraints imposed on them. These constraints are enforced both by JavaScript on the client side and by the server.

- (i) *Job name*: a short descriptive name for the job. Only alphanumeric characters are allowed. If no name is entered, the job's unique tag (based on the time of submission) is used. The job name is used in the title of the output page and printed on each plot.
- (ii) *Sequences*: two different sequences should be entered using the letters A, C, G, T, U and N. (Case is irrelevant.) T and U are considered equivalent (whether to interpret the sequences as RNA or DNA is specified with a different field), and N indicates an unknown base. We currently do not support the IUPAC ambiguous symbols R, Y, S, W, K, M, B, D, H and V; each of these is converted to an N, as are other alphabetic characters. Non-alphabetic characters are discarded. The server currently enforces a maximum length of 50 bases for each sequence.
- (iii) *Temperature range*: the minimum and maximum temperature T_{\min} and T_{\max} for simulation, as well as the temperature increment T_{inc} , in $^{\circ}\text{C}$. The calculations are performed at T_{\min} , $T_{\min} + T_{\text{inc}}$, \dots , so the final value may not be exactly T_{\max} . The number of temperatures in the range

$$1 + \left\lceil \frac{T_{\max} - T_{\min}}{T_{\text{inc}}} \right\rceil$$

may not exceed 200.

- (iv) *Nucleic acid type*: whether to interpret the sequence as RNA or DNA. The server uses the latest energy rules in each case.
- (v) *Initial concentrations*: the strand concentrations for each sequence, in mol/l (M). Naturally, both concentrations must be positive.
- (vi) *Salt conditions*: the concentrations of sodium and magnesium ions, in mol/l (M) or mmol/l (mM). In the default 'oligomer' mode, $[\text{Na}^+]$ must be between 0.01 and 1 M, and $[\text{Mg}^{2+}]$ must be <0.1 M. The alternative 'polymer mode', better suited for structures with stems of >20 bp, allows changing $[\text{Na}^+]$ only. Salt conditions apply only to DNA.
- (vii) *Email address*: if a valid email address is entered, the user will be notified when the job is ready.

Output

Each job produces a variety of output in both textual and graphical forms.

First, a simple form allows the user to display a plot of base pair probabilities for any species at any temperature. A grid is displayed with the color and size of the dot at position (i, j) indicating the conditional probability of base i pairing with base j given that at least 1 bp forms.

Second, several plots are displayed, hyperlinked to larger versions. Each plot is also available for download as PostScript or PDF.

- (i) *Concentration plot*: the relative concentrations of each of the five species (one heterodimer, two homodimers and two single strands) in the ensemble is plotted as a function of temperature, with the single strands subdivided into folded and unfolded states. The seven curves sum to one at each temperature. The text file from which the concentration plot is generated is also available for download.

- (ii) *Heat capacity plot*: the heat capacity of the ensemble, computed by numerical differentiation of the ensemble free energy, is plotted as a function of temperature. The maximum value is identified and labeled with the melting temperature $T_m(C_p)$. A second plot is also available (though not displayed) that shows the contributions of each species to the ensemble heat capacity.
- (iii) *Absorbance plot*: the expected UV absorption is plotted as a function of temperature. Since UV absorption is essentially a measure of the number of base pairs present, there are two ways to determine a melting temperature from this curve. Either the inflection point or the point halfway between the minimum and the maximum values may be taken as $T_m(\text{Ext})$; we use the latter. As with the concentration plot, the text file containing the raw data that was plotted can be downloaded as well. As with heat capacity, a second absorbance plot is available that shows the contributions of each species to the ensemble absorbance.

Third, several text files containing thermodynamic data are available. Files containing the free energy, heat capacity, enthalpy and entropy at each temperature are available for the ensemble.

Finally, the user can download the entire job as one file, either in zip format or as a tar archive compressed with gzip or bzip2.

Figure 1 shows a sample concentration plot, and Figure 2 shows sample heat capacity and absorbance curves. The Supplementary Material contains examples of other types of plots produced by the server. We added a plot comparing simulated UV absorbance with measured UV absorbance and another comparing simulated heat capacity with corresponding measured values.

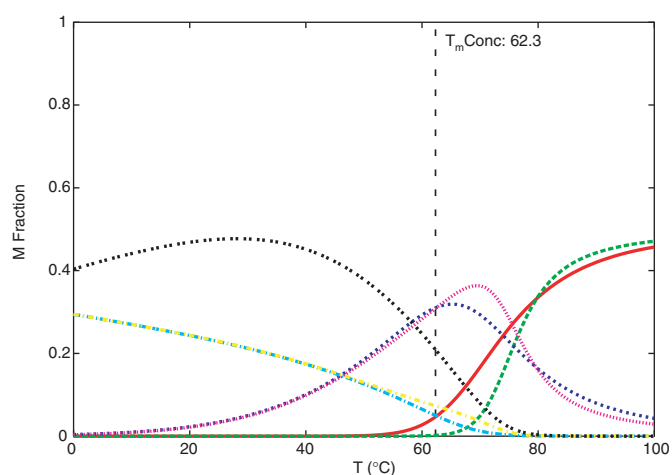


Figure 1. A typical concentration plot, resulting from the simulation of 10 μM of each of A = 5'-GTGTTTATATACTGCGGCAGTATGTAGACAC-3' and B = 5'-GTGTTTATATACTGCTGCAGTATAAACAC-3' with $[\text{Na}^+] = 1$ M and $[\text{Mg}^{2+}] = 0$ M. The mole fraction of each species is plotted as a function of temperature. The red and green lines indicate the concentrations of the unfolded single strands, and the blue and magenta lines show the folded single strands. The yellow and cyan curves correspond to the two homodimers and the black curve to the heterodimer.

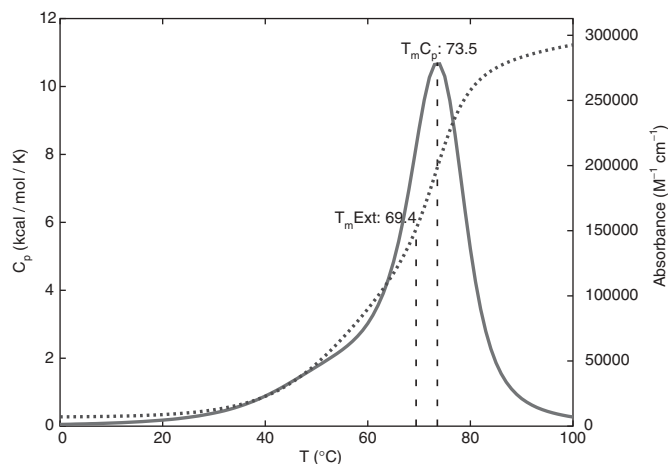


Figure 2. Heat capacity and absorbance curves for the example from Figure 1. The heat capacity (left axis label) is plotted with a solid line, while the absorbance (right axis label) is plotted with a dotted line.

Advanced form

In addition to the simple input form described above, an advanced form is also available. A hyperlink at the top of the page allows the user to switch between the simple and the advanced forms; this preference is saved as an HTTP cookie.

The advanced form contains several options not present in the simple form:

- (i) *Program*: by default the server uses the hybrid2 program, which computes a partition function for each species. However, the advanced user may instead choose to use hybrid2-min, which calculates a minimum energy and corresponding structure for each species.
- (ii) *Advanced options*: by default, a prefilter and a postfilter are enabled that reduce the number of spurious structures counted; the advanced user may choose to disable these filters. Also, the user may elect to skip computation of probabilities; this results in a significantly faster computation at the expense of the probability and absorbance plots.
- (iii) *Exclude species*: to save time, the advanced user may choose to exclude one or more species from consideration. Each species except the heterodimer can be individually excluded, allowing any subset of the five species containing the heterodimer to be chosen.
- (iv) *Enthalpy/entropy for single strands*: by default, the DINAMelt software assigns to the unfolded single strands an enthalpy equal to 10% of the ensemble enthalpy. The entropy for these unfolded single strands is chosen to obtain a melting temperature of 50°C, i.e.

$$\Delta S_{ss} = \frac{1000 \times \Delta H_{ss}}{T_{melt} + 273.15},$$

where ΔS_{ss} is expressed in e.u., ΔH_{ss} in kcal/mol and T_{melt} in °C. The advanced user may choose different values for the fraction and the melting temperature.

Single sequence

The five species method described above requires two different sequences. If the sequences are the same, or equivalently

there is only one sequence, then the number of species is reduced to two: one homodimer and one single strand. A separate form, also with simple and advanced versions, is available for this case. This form contains only one sequence input and one strand concentration.

EQUIPMENT AND ORGANIZATION

The current web server is running on equipment donated to Rensselaer Polytechnic Institute (RPI) by IBM Research in the fall of 2001. The server consists of 36 nodes, each with dual 1 GHz Intel Pentium III processors and 1 GB of RAM. The operating system is Red Hat Linux 7.3. All equipment was originally assembled and housed at the Academy of Electronic Media (<http://www.academy.rpi.edu/>) and moved to the Voorhees Computing Center in March 2003.

FUTURE DEVELOPMENT

The DINAMelt web server currently offers stable versions of software developed as part of a continuing research program. We intend to update the server as new or improved methods become available. Work is already in progress on two projects.

The current software does not allow intramolecular base pairs in hybridized species. That is, if A and B hybridize, each base pair links a nucleotide in A with one in B. A new hybridization program under development will allow both intermolecular and intramolecular base pairs.

The values of ΔH_{ss} and ΔS_{ss} are currently chosen by an empirically derived *ad hoc* rule. IDT has provided us with some preliminary UV and C_p melting profiles for single-stranded, unfolded DNA. The next step will be to test models that consider the dinucleotide compositions of both strands. It is already clear, for example, that ΔH_{ss} (per dinucleotide) is about -1.5 kcal/mol for poly(dC) and 0 kcal/mol for poly(dT).

CITING THE DINAMelt WEB SERVER

Authors who make use of the DINAMelt web server should cite this article as a general reference and should also include the URL to the entrance page, <http://www.bioinfo.rpi.edu/applications/hybrid/>. The web server pages will list additional articles for citation that relate to the algorithms employed, the software that implements them and the energy parameters it uses.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank Art Sanderson, former Vice President of Research at RPI, for connecting us with the Academy of Electronic Media and for supporting this project; Bill Shumway, for initiating and facilitating interactions with IBM Research; and Alex Yu, who has done so much work in assembling the hardware and in keeping the server running day in and day out. This work was supported, in part, by grant no. GM54250 to M.Z. from

the National Institutes of Health and by a Graduate Fellowship to N.R.M. from RPI. Finally, we thank IBM Research for the SUR grant that gave us a very powerful resource for offering this valuable service. Funding to pay the Open Access publication charges for this article was provided by private RPI funds.

Conflict of interest statement. None declared.

REFERENCES

1. Rouillard, J.-M., Herbert, C.J. and Zuker, M. (2002) OligoArray: genome-scale oligonucleotide design for microarrays. *Bioinformatics*, **18**, 486–487.
2. Tyagi, S. and Kramer, F.R. (1998) Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.*, **14**, 303–308.
3. SantaLucia, J., Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
4. Walter, A.E., Turner, D.H., Kim, J., Lyttle, M.H., Müller, P., Mathews, D.H. and Zuker, M. (1994) Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl Acad. Sci. USA*, **91**, 9218–9222.
5. Dimitrov, R.A. and Zuker, M. (2004) Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys. J.*, **87**, 215–226.
6. Puglisi, J.D. and Tinoco, I., Jr (1989) Absorbance melting curves of RNA. *Methods Enzymol.*, **180**, 304–325.
7. Owczarzy, R., You, Y., Moreira, B.G., Manthey, J.A., Huang, L., Behlke, M.A. and Walder, J.A. (2004) Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry*, **43**, 3537–3554.